ECE 509: Convex Optimization **Rutgers: Spring 2024**

Homework #4

February 29, 2024

Name: Ravi Raghavan *Extension: No*

Problem 1. *Initial point and sublevel set condition.* Consider the function $f(x) = x_1^2 + x_2^2$ with domain $dom f = \{(x_1, x_2) | x_1 > 1\}$.

(a) What is p^* and is it attained by any $x \in dom(f)$?

Solution. $p^* = \inf_x f(x) = \lim_{x \to (1,0)} f(x) = 1$. Unfortunately, this is NOT attained by any $x \in dom(f)$

(b) Draw the sublevel set $S = \{x | f(x) \leq f(x^{(0)})\}$ for $x^{(0)} = (2, 2)$. Is the sublevel set S closed? Is f strongly convex on S?

Solution.

Note: The sketch is on the next page

This sublevel set is NOT closed. Let's look at the sequence $(\frac{t+1}{t}, 1)$ for $t \in \mathbb{N}$. Please note that, in this definition, the Natural Numbers are $1, 2, 3, ...$ It DOES NOT include 0. This sequence is contained within the sublevel set but, the limit point of $(1, 1)$ is NOT in the sublevel set

f IS strongly convex on S. $\nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$, $\nabla^2 f(x) = \begin{pmatrix} 2 & 0 \\ 0 & 2 \end{pmatrix}$. Hence, we can clearly find an m and M such that $mI \preceq \nabla^2 f(x) \preceq MI$

(c) What happens if we apply the gradient method with backtracking line search, starting at $x^{(0)}$? Does $f(x^{(k)})$ converge to p^*?

Solution. We can see that $\nabla f(x) = \begin{pmatrix} 2x_1 \\ 2x_2 \end{pmatrix}$, $-\nabla f(x) = \begin{pmatrix} -2x_1 \\ -2x_2 \end{pmatrix}$

We start our iteration at $(2, 2)$. When $x_1 = x_2$, it is clear to see that $\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2}$. Let $x^{(k)}$ represent iterate at the end of iteration k.

Hence, we know that $x^{(k+1)} = x^{(k)} + t\Delta x^{(k)}$. Since we are doing the gradient method with backtracking line search, we can further say that $x^{(k+1)} = x^{(k)} - t\nabla f(x^{(k)})$

Since, for $x^{(0)}$, $x_1 = x_2$, and we know that when $x_1 = x_2$, $\frac{\partial f}{\partial x_1} = \frac{\partial f}{\partial x_2}$, we can say that for $x^{(k+1)}$, x_1 will equal x_2.

Hence, it is clear to see that our descent is continuing along the line $x_1 = x_2$. It is also clear to see that the iterate sequence is converging to $(1, 1)$. We obviously cannot reach $(1, 1)$ or go beyond $(1, 1)$ since we would then be OUTSIDE the function's domain.

We can state that $\lim_{k \to \infty} x^{(k)} = (1, 1)$ and $\lim_{k \to \infty} f(x^{(k)}) = 2$. It is clear to see that $f(x^{(k)})$ does NOT converge to p^*

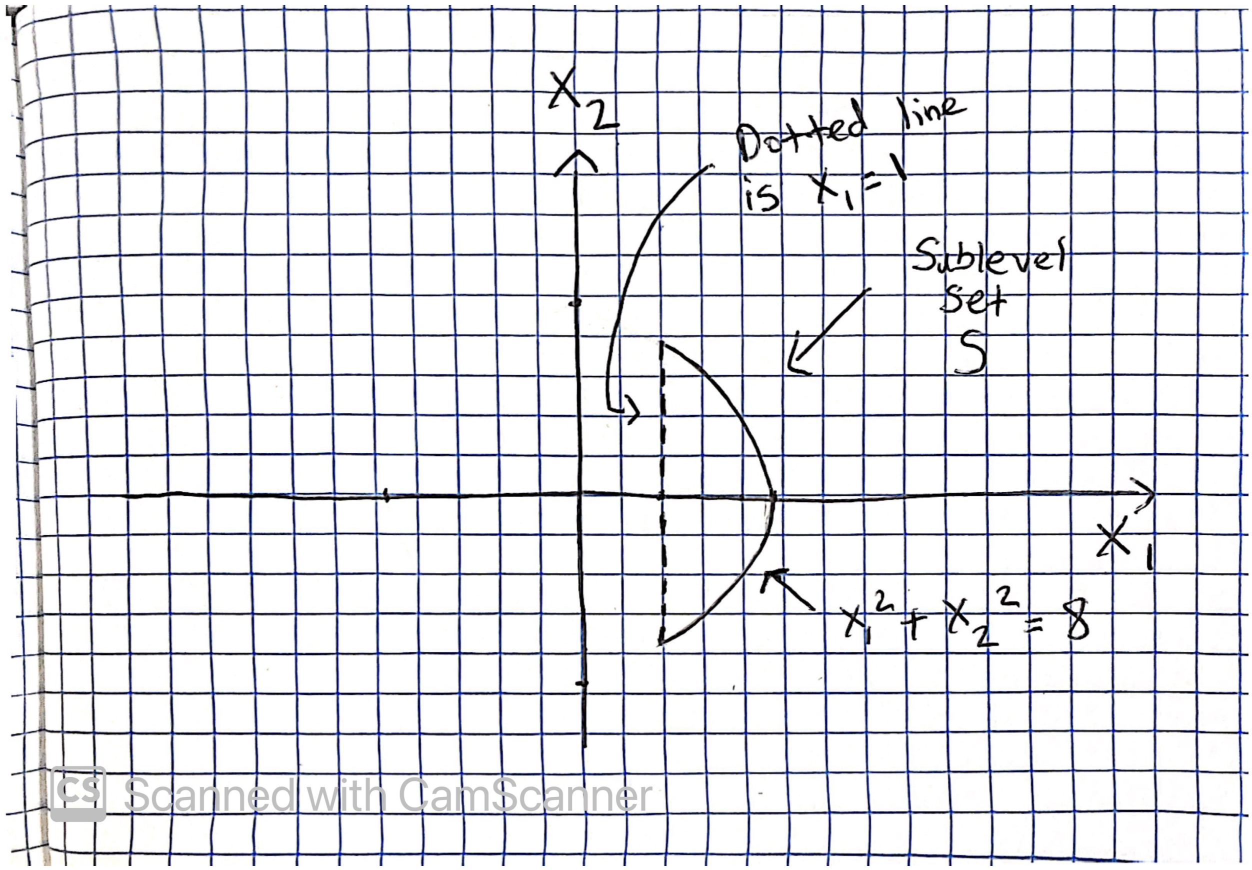

Figure 1: Sketch of Sublevel Set

Problem 2. *Backtracking Line Search* Suppose f is strongly convex with $mI \preceq \nabla^2 f(x) \preceq MI$. Let Δx be a descent direction at x. Show that the backtracking stopping condition holds for

$$0 < t \leq -\frac{\nabla f(x)^T \Delta x}{M||\Delta x||_2^2} \tag{1}$$

Use this to give an upper bound on the number of backtracking iterations.

Solution. Let's briefly revisit the backtracking algorithm

Algorithm 1 Backtracking

Parameters: Δx(Given Descent Direction for f at $x \in dom f$) , $\alpha \in (0, 0.5)$, $\beta \in (0, 1)$
$t = 1$
while $f(x + t\Delta x) > f(x) + \alpha t \nabla f(x)^T \Delta x$ **do**
  $t = \beta t$
end while

Based on the definition of strong convexity, we know that

$$f(y) \leq f(x) + \nabla f(x)^T (y - x) + \frac{M}{2}||y - x||_2^2 \tag{2}$$

Hence, if we set $y = x + t\Delta x$

$$f(x+t\Delta x) \le f(x) + \nabla f(x)^T(x+t\Delta x - x) + \frac{M}{2}||x+t\Delta x - x||_2^2 \tag{3}$$

$$f(x+t\Delta x) \le f(x) + t\nabla f(x)^T(\Delta x) + \frac{Mt^2}{2}||\Delta x||_2^2 \tag{4}$$

Backtracking Stopping Condition: $f(x+t\Delta x) \le f(x) + \alpha t\nabla f(x)^T(\Delta x)$
Due to the definition of strong convexity, we are guaranteed for this backtracking stopping condition to hold when:
$f(x) + \alpha t\nabla f(x)^T(\Delta x) \ge f(x) + t\nabla f(x)^T(\Delta x) + \frac{Mt^2}{2}||\Delta x||_2^2$

$\alpha t\nabla f(x)^T(\Delta x) \ge t\nabla f(x)^T(\Delta x) + \frac{Mt^2}{2}||\Delta x||_2^2$

$0 \ge (1-\alpha)t\nabla f(x)^T(\Delta x) + \frac{Mt^2}{2}||\Delta x||_2^2$

$0 \ge t[(1-\alpha)\nabla f(x)^T(\Delta x) + \frac{Mt}{2}||\Delta x||_2^2]$

We must have $t > 0$ AND $(1-\alpha)\nabla f(x)^T(\Delta x) + \frac{Mt}{2}||\Delta x||_2^2 \le 0$

$(1-\alpha)\nabla f(x)^T(\Delta x) + \frac{Mt}{2}||\Delta x||_2^2 \le 0$
$\frac{Mt}{2}||\Delta x||_2^2 \le (\alpha - 1)\nabla f(x)^T(\Delta x)$
$t \le \frac{2(\alpha-1)\nabla f(x)^T(\Delta x)}{M||\Delta x||_2^2}$

Hence, we are guaranteed for the backtracking condition to hold when $t \le \frac{2(\alpha-1)\nabla f(x)^T(\Delta x)}{M||\Delta x||_2^2}$

Since $\alpha \in (0, 0.5)$, we can see that $\frac{2(\alpha-1)\nabla f(x)^T(\Delta x)}{M||\Delta x||_2^2} \ge -\frac{\nabla f(x)^T\Delta x}{M||\Delta x||_2^2}$

Hence, we see that, when $0 < t \le -\frac{\nabla f(x)^T\Delta x}{M||\Delta x||_2^2}$, the backtracking stopping condition holds.

Let $t_1 = \frac{2(\alpha-1)\nabla f(x)^T(\Delta x)}{M||\Delta x||_2^2}$
For an upper bound on the number of backtracking iterations, assuming $t_1 \le 1$, we need:
$\beta^k \le t_1$
$k \ge \log_\beta(t_1)$

Problem 3. *Quadratic problem in* $\mathbb{R}^2$. Verify the expressions for the iterates $x^{(k)}$ in the first example of 9.3.2

Solution. $f(x) = \frac{1}{2}(x_1^2 + \gamma x_2^2)$, $\frac{\partial f}{\partial x_1} = x_1$ and $\frac{\partial f}{\partial x_2} = \gamma x_2$, $\nabla f(x) = (x_1, \gamma x_2)$

We know that, at $k = 0$, $x^0 = (\gamma, 1)$

$x^{(k)} - \alpha\nabla f(x^{(k)}) = \begin{pmatrix} (1-\alpha)x_1 \\ (1-\alpha\gamma)x_2 \end{pmatrix}$

<u>Exact Line Search:</u>
Given the expressions for the iterates $x^{(k)}$, let's conduct exact line search. In the book, the iterate expressions are given as follows:

$x_1^{(k)} = \gamma(\frac{\gamma-1}{\gamma+1})^k$ and $x_2^{(k)} = (-\frac{\gamma-1}{\gamma+1})^k$

The book states that $f(x^{(k)}) = \frac{\gamma(\gamma+1)}{2}(\frac{\gamma-1}{\gamma+1})^{(2k)}$

Substituting these expressions into $x^{(k)} - \alpha\nabla f(x^{(k)})$ gives us:

$$x^{(k)} - \alpha\nabla f(x^{(k)}) = \begin{pmatrix}(1-\alpha)x_1 \\ (1-\alpha\gamma)x_2\end{pmatrix} = (\frac{\gamma-1}{\gamma+1})^k \begin{pmatrix}(1-\alpha)\gamma \\ (1-\alpha\gamma)(-1)^k\end{pmatrix}$$

$$f(x^{(k)} - \alpha\nabla f(x^{(k)})) = \frac{1}{2}(\frac{\gamma-1}{\gamma+1})^{2k}((1-\alpha)^2\gamma^2 + \gamma(1-\alpha\gamma)^2)$$

To minimize $g(\alpha) = f(x^{(k)} - \alpha\nabla f(x^{(k)}))$, it is clear that we must minimize $((1-\alpha)^2\gamma^2 + \gamma(1-\alpha\gamma)^2)$

$\frac{d}{d\alpha}((1-\alpha)^2\gamma^2 + \gamma(1-\alpha\gamma)^2) = -2(1-\alpha)\gamma^2 - 2\gamma^2(1-\alpha\gamma)$

Setting this derivative equal to 0 gives us:
$-2(1-\alpha)\gamma^2 - 2\gamma^2(1-\alpha\gamma) = 0$
$-2\gamma^2 + 2\alpha\gamma^2 - 2\gamma^2 + 2\alpha\gamma^3 = 0$

Divide this entire equation by $2\gamma^2$
$-1 + \alpha - 1 + \alpha\gamma = 0$
$\alpha + \alpha\gamma = 2$
$\alpha = \frac{2}{1+\gamma}$

Substituting the value of alpha, we get:

$$x^{(k+1)} = x^{(k)} - \alpha\nabla f(x^{(k)}) = \begin{pmatrix}(1-\alpha)x_1^{(k)} \\ (1-\alpha\gamma)x_2^{(k)}\end{pmatrix} = \begin{pmatrix}\frac{\gamma-1}{\gamma+1}x_1^{(k)} \\ \frac{1-\gamma}{\gamma+1}x_2^{(k)}\end{pmatrix}$$

<u>**Verification:**</u>
First, we show that the closed form expressions are TRUE for $k = 0$
$x_1^{(0)} = \gamma = \gamma(\frac{\gamma-1}{\gamma+1})^0$, $x_2^{(0)} = 1 = (-\frac{\gamma-1}{\gamma+1})^0$, $f(x^{(0)}) = \frac{1}{2}(\gamma^2+\gamma) = \frac{\gamma(\gamma+1)}{2} = \frac{\gamma(\gamma+1)}{2}(\frac{\gamma-1}{\gamma+1})^{(2*0)} = \frac{\gamma(\gamma+1)}{2}(\frac{\gamma-1}{\gamma+1})^0$

Now, assuming the closed form expressions are TRUE for k, show that they must be true for $k+1$

$$x^{(k+1)} = x^{(k)} - \alpha\nabla f(x^{(k)}) = \begin{pmatrix}\frac{\gamma-1}{\gamma+1}x_1^{(k)} \\ \frac{1-\gamma}{\gamma+1}x_2^{(k)}\end{pmatrix} = \frac{\gamma-1}{\gamma+1}\begin{pmatrix}x_1^{(k)} \\ -x_2^{(k)}\end{pmatrix}$$

$$x_1^{(k+1)} = \frac{\gamma-1}{\gamma+1}x_1^{(k)} = \frac{\gamma-1}{\gamma+1}(\gamma(\frac{\gamma-1}{\gamma+1})^k) = \gamma(\frac{\gamma-1}{\gamma+1})^{(k+1)}$$

$$x_2^{(k+1)} = \frac{\gamma-1}{\gamma+1}(-x_2^{(k)}) = -\frac{\gamma-1}{\gamma+1}(-\frac{\gamma-1}{\gamma+1})^k = (-\frac{\gamma-1}{\gamma+1})^{(k+1)}$$

$$f(x^{(k+1)}) = \frac{1}{2}(\frac{\gamma-1}{\gamma+1})^{2(k+1)}(\gamma^2+\gamma) = \frac{\gamma(\gamma+1)}{2}(\frac{\gamma-1}{\gamma+1})^{2(k+1)}$$

This completes the verification and shows that the following are TRUE:
$x_1^{(k)} = \gamma(\frac{\gamma-1}{\gamma+1})^{(k)}$
$x_2^{(k)} = (-\frac{\gamma-1}{\gamma+1})^{(k)}$
$f(x^{(k)}) = \frac{\gamma(\gamma+1)}{2}(\frac{\gamma-1}{\gamma+1})^{2k}$
